



Al-Mualla, MES., Canagarajah, CN., & Bull, DR. (2000). On the performance of temporal error concealment for long-term motion-compensated prediction. In *IEEE Int. Conf. on Image Processing (ICIP), Vancouver* (Vol. 3, pp. 376 - 379). Institute of Electrical and Electronics Engineers (IEEE).
<https://doi.org/10.1109/ICIP.2000.899407>

Peer reviewed version

Link to published version (if available):
[10.1109/ICIP.2000.899407](https://doi.org/10.1109/ICIP.2000.899407)

[Link to publication record in Explore Bristol Research](#)
PDF-document

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

ON THE PERFORMANCE OF TEMPORAL ERROR CONCEALMENT FOR LONG-TERM MOTION-COMPENSATED PREDICTION

Mohammed E. Al-Mualla, C. Nishan Canagarajah and David R. Bull

Image Communications Group, Centre for Communications Research, University of Bristol,
Merchant Venturers Building, Woodland Road, Bristol BS8 1UB, United Kingdom,
Tel: +44-117-9545126, Fax: +44-117-9545206,
e-mail: {m.almualla, nishan.canagarajah, dave.bull}@bristol.ac.uk

ABSTRACT

This paper investigates the performance of different temporal concealment techniques when incorporated within a long-term motion compensated video codec. In particular, the paper finds a combination of techniques that best recovers the spatial-temporal components of a damaged long-term motion vector. It is demonstrated also that accurate recovery of the spatial components of the damaged motion vector is, in general, more important and more complex than the recovery of the temporal component. Such findings are explained in view of the properties of the long-term block motion field. The paper, also, proposes a novel long-term multihypothesis temporal concealment technique where the damaged block is concealed using the average of a number of candidates from the several reference frames available at the decoder. The superior performance of this technique is demonstrated over a range of block error rates.

1. INTRODUCTION

Fast growing multimedia services have generated a great interest in transmitting digital video over a wide range of communication channels. However, practical communication channels are not error free and the quality of the reconstructed video can be severely degraded.

One method to combat the effects of channel errors is *error concealment* which provides a subjectively acceptable approximation to the original data by exploiting the high *temporal* and/or *spatial* correlation of video sequences. Unlike other methods, error concealment, in general, does not increase the bit rate, requires no change to the encoder, introduces no delays and can be applied in almost any application. This makes it particularly attractive for very low bit rate real-time applications like wireless video.

An important class of error concealment methods is *temporal concealment* techniques. Such techniques exploit the

high temporal correlation of video signals and conceal damaged blocks using information from correctly received and/or previously concealed pels in a reference frame. Temporal concealment can be viewed as a *motion information recovery* process where the decoder attempts to *recover* an approximation, $\hat{\mathbf{d}}$, of the damaged motion vector, $\mathbf{d} = (d_x, d_y)$. Such process is normally based on certain properties of the block motion fields of typical video sequences.

Recently, *long-term motion estimation* has been proposed to improve the coding efficiency of video codecs [1, 2]. In this case, the motion vector, \mathbf{d} , is extended by a *temporal component*, d_t , permitting the use of several reference frames for motion estimation and compensation.

A number of temporal concealment techniques have been proposed in the literature [3, 4, 5, 6], and their performances have been extensively studied within typical *single-reference* video codecs operating over various error-prone channels. There is, however, a *need* to characterise the performance of such techniques within *long-term* motion-compensated video codecs. In particular, what is the best method to recover the new added temporal component? Are the properties of the long-term block motion field different from those of the single-reference field? and how does this influence the temporal concealment process?

In this paper we investigate the performance of different temporal concealment techniques when included within a long-term motion compensated video codec. In particular we find a combination of techniques that best recovers the spatial-temporal components of a damaged long-term motion vector. We also compare the importance and the complexity of the two processes: spatial components recovery and temporal component recovery. We explain all our findings in view of the properties of the long-term block motion field. In addition, we propose a novel long-term temporal concealment technique where the damaged block is concealed using the average of a number of candidates from the several reference frames available at the decoder.

This work was supported by Emirates Telecommunications Corporation (ETISALAT), United Arab Emirates.

2. TEMPORAL CONCEALMENT TECHNIQUES

In this paper, four temporal concealment techniques are considered:

ZR The recovered motion component (either spatial or temporal) is set to *zero*. Thus, $\hat{d} = 0$. Note that a temporal component of $d_t = 0$ refers to the most recent frame in the multi-frame memory.

AV The recovered motion component is the *average* of the corresponding components of a set of neighbouring motion vectors. Thus, for 4 neighbouring vectors:

$$\hat{d} = 1/4(d_L + d_R + d_T + d_B),$$

where L, R, T , and B stands for left, right, top and bottom neighbours, respectively.

BM This is a *boundary matching* method [3]. A set of candidate motion vectors $\{d_i\}$ is first chosen. Each candidate is then used to conceal the damaged block. The quality of this concealment can be assessed using the *side match distortion* (SMD) measure which is defined as the sum of absolute (or squared) differences across the 4 boundaries of the block. Thus:

$$\hat{d} = \arg \min_{d_i} \text{SMD}(d_i).$$

MI This method was introduced by the authors in [4, 6]. Conventional methods (like ZR, AV and BM) recover one vector, \hat{d} , for the whole damaged block. MI, however, uses *motion field interpolation* to recover one vector, $\hat{d}(x, y)$, per pel of the damaged block. That is, each pel is concealed individually. Thus:

$$\hat{d}(x, y) = 1/2((1-x_n)d_L + x_nd_R + (1-y_n)d_T + y_nd_B),$$

where (x_n, y_n) are the normalised spatial co-ordinates of the pel within the block.

3. TEMPORAL CONCEALMENT FOR LONG-TERM MOTION-COMPENSATED PREDICTION

In this section we investigate the performance of different combinations of the above temporal concealment techniques when included within a long-term motion compensated video codec. Each combination is of the form S-T, where S is the technique (from the list in section 2) used to recover the spatial components (d_x, d_y) , whereas T is the technique (from the same list) used to recover the temporal component d_t . Since there are 4 techniques in the list, there are 16 possible combinations. Because of space limitations we only show here a number of them. The long-term video

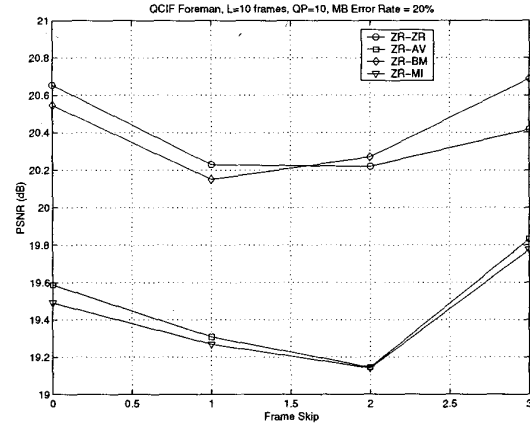


Fig. 1. Temporal component recovery

codec used in our simulations is an H.263-like codec which utilises $L = 10$ reference frames for motion estimation and compensation. Thus, $0 \leq d_t \leq 9$ where $d_t = 0$ refers to the most recent reference frame in memory. Motion is estimated using full-search full-pel block matching, restricted motion vectors in the range ± 15 , 16×16 blocks and SAD as the distortion measure. Motion estimation and mode selection are based on the high complexity mode of the TMN10 H.263 test model [7]. All quoted PSNRs are for the luminance component only.

3.1. Temporal Component Recovery

In the first set of experiments we investigate the best technique for recovering the temporal component. Figure 1 shows the performance of the codec with a quantisation parameter of $QP = 10$ at different frame skips when 20% of the macroblocks are randomly damaged per frame. In this case, S is kept constant at ZR whereas T is changed over ZR, AV, BM and MI.

The best temporal component recovery is achieved by ZR and BM. The good performance of ZR is due to the fact that $d_t = 0$ is the most probable temporal component within the long-term block motion field, as illustrated in Figure 2. Note that at low frame skips (i.e. high frame rates), the simple ZR method is sufficient, whereas at high frame skips the more complex method of BM has to be used. This may be due to the fact that at higher frame skips $d_t = 0$ becomes less probable than in low frame skips, as shown in Figure 2. Thus, at high frame skips longer temporal components start to appear more frequently in the motion field and they need to be recovered using BM. Both AV and MI provide poor temporal component recovery compared to BM and ZR.

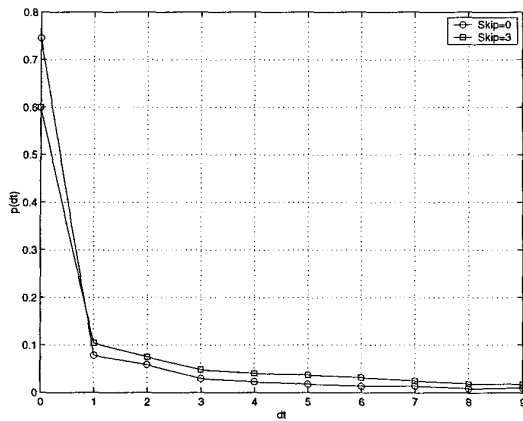


Fig. 2. Frequency of occurrence of the temporal component d_t

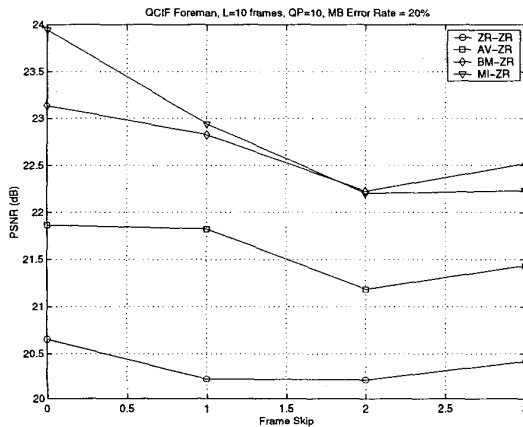


Fig. 3. Spatial components recovery

3.2. Spatial Components Recovery

In the second set of experiments we investigate the best technique for recovering the spatial components. Figure 3 was generated using the same simulation conditions of Figure 1 but in this case T was kept constant at ZR, whereas S was changed over ZR, AV, BM and MI.

The MI technique provides the best spatial component recovery followed closely by BM with ZR giving the worst performance. This is similar to the single-reference results reported by the authors in [4, 6]. Thus, moving from a single-reference system to a multiple-reference system does not significantly influence the spatial component recovery process. A very interesting point to note is that at high frame skips, MI becomes slightly inferior to BM. This may be due to the fact that at high frame skips, the spatial components within the motion field become less correlated, as

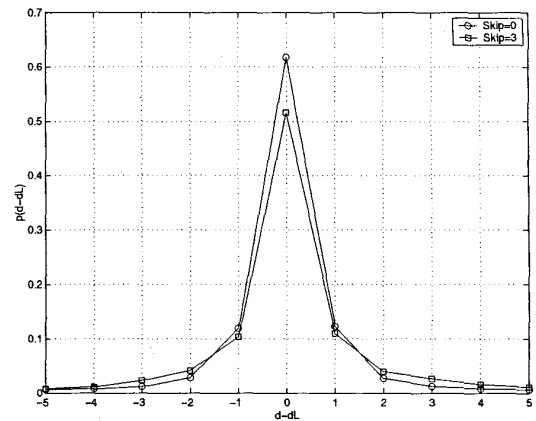


Fig. 4. Correlation between spatial component of a vector and its left neighbour

illustrated in Figure 4. Since MI assumes high correlation between the spatial components, its performance will deteriorate with decreased correlation.

3.3. Spatial-Temporal Components Recovery

Comparing figures 1 and 3 shows that, in general, spatial components recovery is more important and more complex than temporal component recovery. In Figure 1 moving from the best technique ZR-BM to the worst technique ZR-MI drops the quality by about 1 dB, whereas in Figure 3 moving from the best technique MI-ZR to the worst technique ZR-ZR can drop the quality by as high as 3 dBs. With temporal component recovery, a simple technique such as ZR can be sufficient, whereas with spatial component recovery more complex techniques like BM and MI are essential. In general, the combination MI-BM provides the best spatial-temporal recovery.

4. MULTIHYPOTHESIS CONCEALMENT FOR LONG-TERM MOTION COMPENSATED PREDICTION

In [4, 6] we have shown that a more robust performance can be achieved if the concealed block is a weighted average of a number of candidate concealments. Thus, in this case each pel is concealed using a weighted average of a number of candidate pels. This is very similar to multihypothesis motion compensation [8], thus we call it *multihypothesis temporal concealment*.

In this paper we propose a multihypothesis temporal concealment technique to be used with long-term motion compensated prediction. In this technique the spatial components are first recovered using MI (as suggested by the results in 3.2). Instead of recovering a single temporal

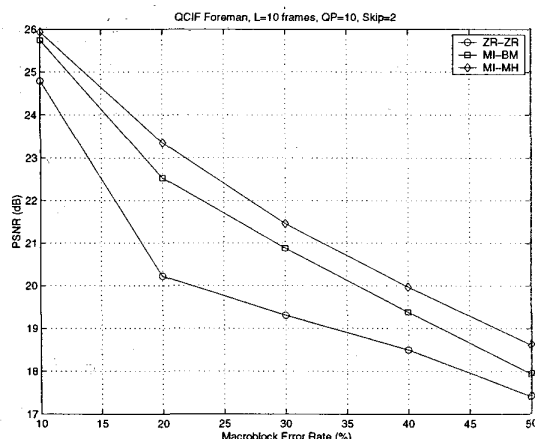


Fig. 5. Comparison between MI-MH, MI-BM and ZR-ZR

component, all 4 neighbouring temporal components are utilised and the block is concealed as the average of the 4 spatially displaced blocks in the corresponding reference frames. Thus

$$\hat{I}_c(x, y) = \sum_{i=1}^4 w_i(x, y) I_r(x + \hat{d}_x(x, y), y + \hat{d}_y(x, y), d_{t_i}),$$

where $\hat{I}_c(x, y)$ refers to the concealed pel in the current frame, $I_r(x, y, d_t)$ refers to a reference frame in the frame-memory with index d_t , $(\hat{d}_x(x, y), \hat{d}_y(x, y))$ are the spatial components recovered, using MI, at pel (x, y) , d_{t_i} are the temporal components of the 4 neighbouring vectors, and $w_i(x, y)$ are the multihypothesis concealment weights. We designate this approach as MI-MH.

In this paper we simply set $w_i(x, y) = 1/4$. However, in [6] we have shown that some improvement can be achieved if the multihypothesis concealment weights are adjusted based on knowledge of the problem at hand. A technique with optimal weights for long-term error concealment is under development.

The performance of the proposed technique over a range of block error rates is illustrated in Figure 5. This figure compares the performance of MI-MH to that of MI-BM (which is the best combination as suggested by subsection 3.3) and also to the performance of ZR-ZR (which is the simplest and most commonly used combination). The superior performance of the proposed technique is immediately evident.

5. CONCLUSIONS

The performance of different temporal concealment techniques when incorporated within a long-term motion compensated video codec was investigated. It was found that

accurate recovery of the spatial components of the damaged motion vector is, in general, more important and must employ more sophisticated techniques than the recovery of the temporal component. A novel long-term multihypothesis temporal concealment technique was proposed. In this technique, the damaged block is concealed using the average of a number of candidates from the several reference frames available at the decoder. The superior performance of this technique was demonstrated over a range of block error rates.

6. REFERENCES

- [1] T. Wiegand, X. Zhang, and B. Girod, "Long-term memory motion-compensated prediction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 1, pp. 70–84, February 1999.
- [2] M. E. Al-Mualla, C. N. Canagarajah, and D. R. Bull, "Simplex minimisation for single- and multiple-reference motion estimation," submitted to *IEEE Transactions on Circuits and Systems for Video Technology*, September 1999.
- [3] W. M. Lam, A. R. Reibman, and B. Liu, "Recovery of lost or erroneously received motion vectors," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1993, vol. 5, pp. 417–420.
- [4] M. E. Al-Mualla, C. N. Canagarajah, and D. R. Bull, "Error concealment using motion field interpolation," in *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, Chicago IL, USA, October 1998, vol. II, pp. 512–516.
- [5] Y. Wang and Q. F. Zhu, "Error control and concealment for video communication: A review," *Proceedings of the IEEE*, vol. 86, no. 5, pp. 974–997, May 1998.
- [6] M. E. Al-Mualla, C. N. Canagarajah, and D. R. Bull, "Motion field interpolation for temporal error concealment," accepted for *IEE-Proceedings, Vision, Image and Signal Processing*, 2000.
- [7] T. R. Gardos, "Video codec test model, near-term, version 10 (TMN10), draft 1," Document Q15-D-65d1, ITU-T, Study Group 16, Tampere, Finland, April 1998, download from <ftp://standard.pictel.com/video-site/h263plus/tmn10.doc>.
- [8] B. Girod, "Efficiency analysis of multihypothesis motion-compensated prediction for video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 2, pp. 173–183, February 2000.